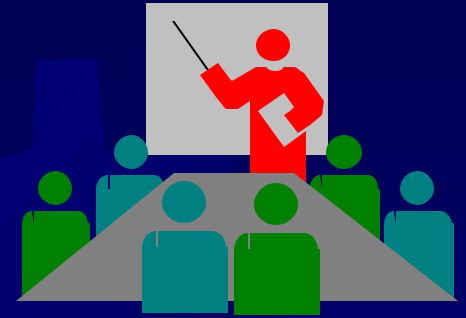


Implications of the Research on Overconfidence for Challenge Problem Solution Strategies



Vicki Bier, University of Wisconsin
August 2002

Algebraic Problem Set



★ $Y = (a + b)^a$:

- a and b are independent, positive, real numbers

★ Example--Problem 2:

- A is contained in the closed interval $A = [a_1, a_2]$
- Information about b is given by n independent sources
- Each source specifies an interval $B_i = [b_1^i, b_2^i]$ of possible values for b
- The B_i can be **consonant** (nested), **consistent**, or **inconsistent**

Questions about Context

- ✱ How were the intervals obtained?
 - ✱ What else (if anything) is known?
- ✱ Are they “infallible” (i.e., contain the true value with probability one)?
 - ✱ Cannot be true if they are inconsistent
- ✱ If the intervals are not infallible:
 - ✱ How likely **are** they to contain the true value?

Questions about Context

- ✱ If intervals are not infallible:
 - ✱ Why does it matter whether the intervals are consonant, consistent, or arbitrary?
- ✱ Emphasizing this suggests replacing consistent intervals by their intersection!
- ✱ What is wrong with that suggestion?
 - ✱ Overconfidence
 - ✱ Dependence

Overconfidence

- ✱ “Assessments can be...*overconfident*, whereby the proportions correct are less than the assessed probabilities”
 - ✱ Lichtenstein et al. (1982)
- ✱ “No problem in judgment and decision making is more prevalent and more potentially catastrophic than overconfidence”
 - ✱ Plous (1993)

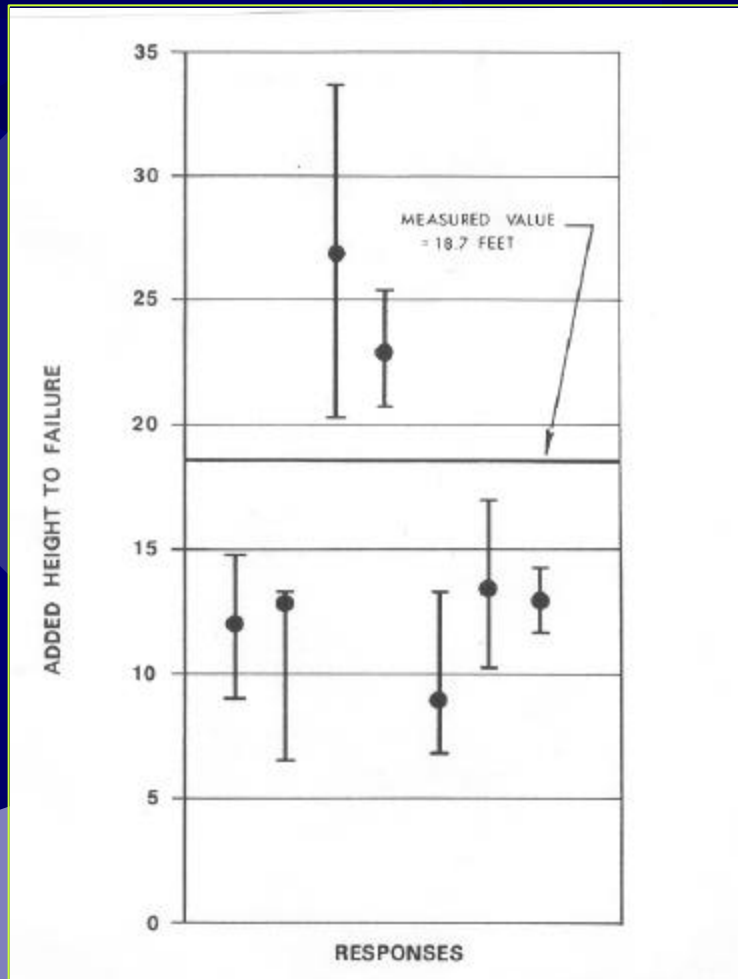
Overconfidence

- ✴ For 98% confidence intervals, the “surprise index”
 - ✴ (Percentage of true values that fall outside the specified range) should be 2%
- ✴ Typical values of the surprise index range from 20% to 40%:
 - ✴ Experts are as overconfident as laypeople!

Overconfidence

- ✱ Thus, intervals obtained from experts are likely to display overconfidence
- ✱ Inconsistent intervals are essentially **guaranteed** to reflect overconfidence!

Example of Overconfidence



- ◆ Height at which an embankment would fail:
 - 7 “internationally known” geo-technical engineers
 - Hynes and Vanmarcke (1976)
- ◆ Error bars fail to contain the true value!
 - Example of “inconsistent” intervals

Examples of Overconfidence

- ☀ Henrion and Fischhoff (1986):

- ☀ Confidence intervals for the velocity of light and other physical constants (gravitational constant, magnetic moment of the proton, etc.)
- ☀ Surprise indices ranged from 0% to 57%

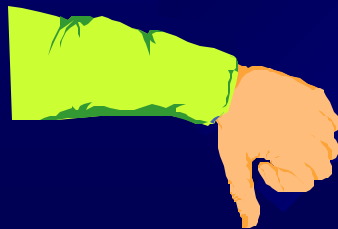
- ☀ Shlyakhter and Kammen (1994):

- ☀ Confidence intervals for 281 elementary particle properties, energy demand forecasts, and population projections
- ☀ Systematic errors of 3σ for population growth (!), 1σ for physical constants and energy demand



Problems with Overconfidence

- ☀ Overconfidence can lead to incorrect decisions:
 - ✱ Accepting excessive risks (if decision maker is risk averse)
 - ✱ Declining desirable risks (if decision maker is risk seeking)
 - ✱ Failing to gather information that could cost-effectively help to reduce uncertainty





Overcoming Overconfidence

- ✱ Cognitive strategies (Plous, 1993):
 - ✱ “Intense performance feedback”
 - ✱ “Stop to consider reasons why your judgment might be wrong”

Overcoming Overconfidence

- ✴ Broadening failure rate distributions:
 - ✴ Some risk analyses treat the stated 5th and 95th percentiles as 20th and 80th percentiles
 - ✴ Martz (1984) suggests treating 5th and 95th percentiles as 12th and 88th percentiles, based on an empirical Bayes model

Overcoming Overconfidence

- ✴ Using long-tailed distributions:

- ✴ E.g., a “compound” distribution, in which a distribution is used to express uncertainty about the extent of overconfidence
- ✴ Shlyakhter (1994), Shlyakhter et al. (1994)

Overcoming Overconfidence

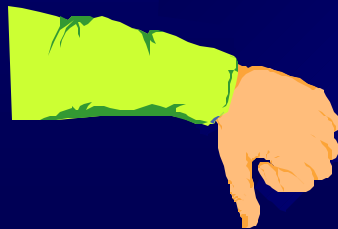
- ✴ Weighting experts by their calibration:
 - ✴ Cooke (1991) suggests weighting experts by how well calibrated and informative they are in a set of empirical calibration questions
 - ✴ This method typically outperforms other methods, and also the best expert, in terms of entropy and calibration on the calibration questions
 - ✴ In some applications (e.g., Cooke, et al., 1994), the majority of experts may be assigned a weight of zero based on poor estimates of known items

Dependence among Experts

- ☀ Expert opinions are also likely to be positively correlated:
 - ✱ May reflect a single school of thought
 - ✱ May reflect “conventional wisdom” in a field
 - ✱ May be influenced by the same data
 - ✱ May be influenced by a single vocal expert
 - ✱ See for example Booker and Meyer (1985)

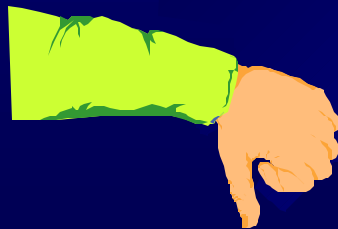
Problems with Dependence

- ✱ Positive correlation among experts can **exacerbate overconfidence**:
 - ✱ If experts are treated as independent, overlap in their intervals will be taken as stronger evidence than is justified
 - ✱ This effect is particularly strong when correlations are large—e.g., > 0.8
 - ✱ See Winkler and Clemen (1992)



Problems with Dependence

- ✱ The value of additional experts decreases rapidly with the number of experts:
 - ✱ Even for small correlations, there may be little additional value provided by consulting more than four or five experts
 - ✱ Infinitely many experts with correlation 0.25 is equivalent to only four independent experts!
 - ✱ See Clemen and Winkler (1985)



Analyzing Dependence

- ✱ Bayesian updating using copulas or other multivariate likelihood functions:
 - ✱ Multivariate normal or lognormal (Winkler, 1981; Chhibber and Apostolakis, 1993)
 - ✱ Dirichlet (Mendel and Sheridan, 1989)
 - ✱ Copulas (Jouini and Clemen, 1996)
- ✱ Other joint distributions can be used

Analyzing Dependence

✴ Clemen et al. (2000):

- ✴ “The most accurate way to obtain a subjective dependence measure is simply to ask...the correlation”
- ✴ “Accuracy can be improved in two ways”—training, and averaging several different dependence measures

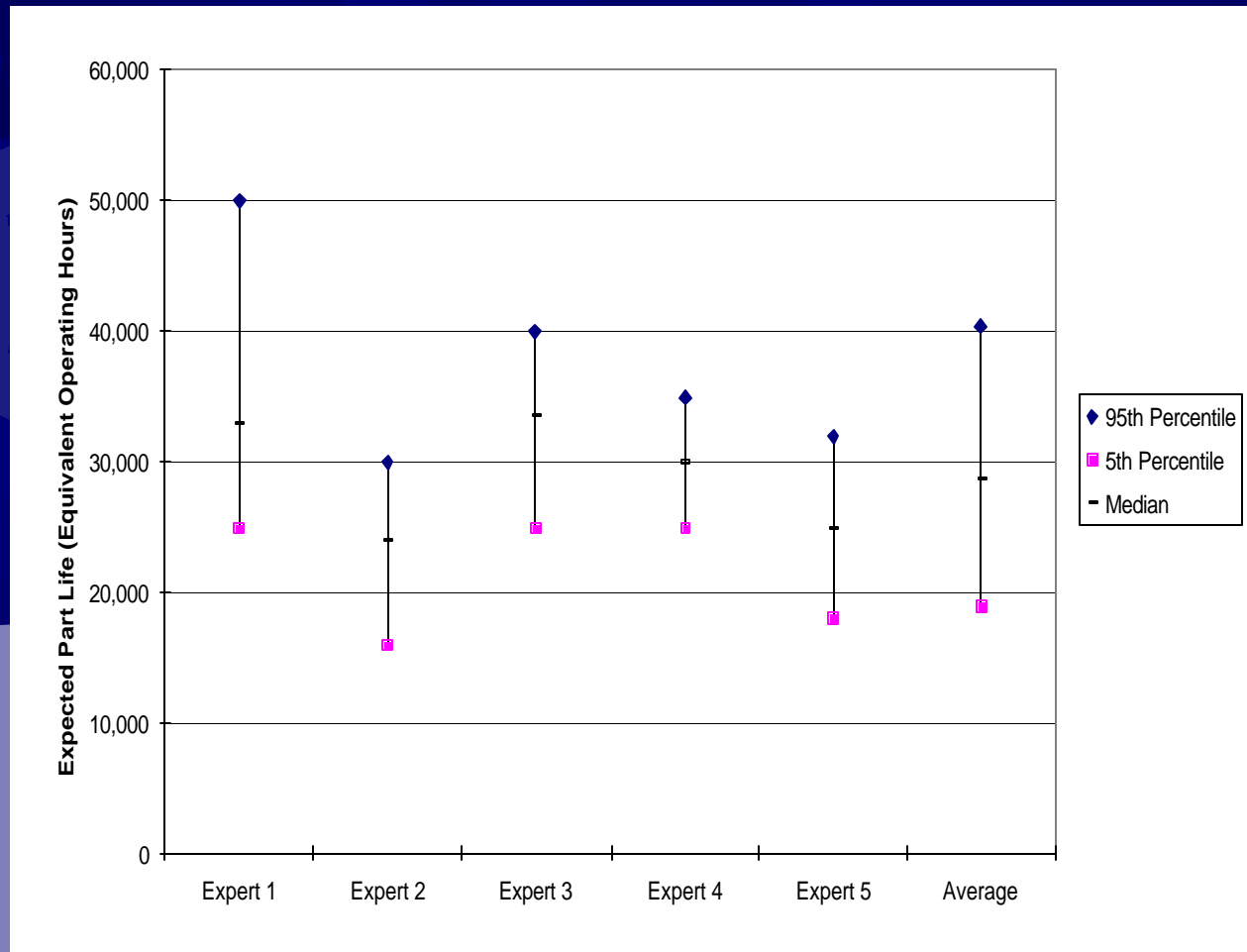
Analyzing Dependence

- ✱ Chhibber and Apostolakis (1993):
 - ✱ “The sensitivity of the...posterior standard deviation to ρ is small”
 - ✱ “Thus, approximations in the assessment of ρ may be acceptable”

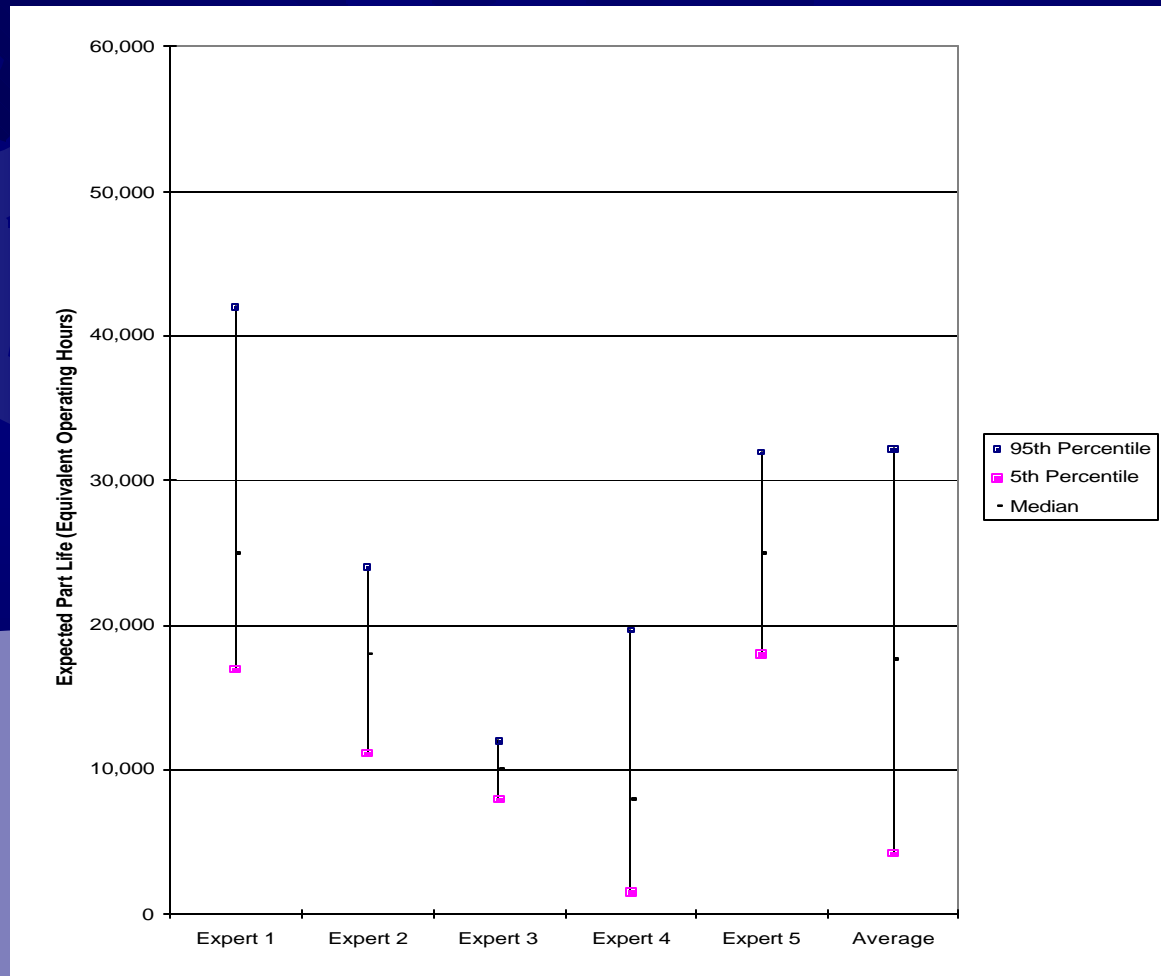
One Simple Strategy

- ✱ Distributions were fit to stated 5th/95th percentiles for expected part life
- ✱ Distributions were combined using arithmetic averaging
 - ✱ (Crystal Ball Monte Carlo software)
- ✱ Each expert was assumed to have an equal probability of being “correct”

Comparison of Estimates (Consistent intervals)



Comparison of Estimates (Inconsistent intervals)



Advantages of Approach

- ✱ Easy to explain
- ✱ Easy to implement
- ✱ No need for judgments of dependence
- ✱ Gives modest weight to “outlier” experts
- ✱ All experts’ opinions are used





Disadvantages of Approach

- ✱ No rigorous theoretical basis
- ✱ Distributions will tend to be:
 - ✱ Too broad when overconfidence and dependence are low
 - ✱ Too narrow when overconfidence and dependence are high
- ✱ May give multi-modal distributions when disagreement among experts is high